Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Association of CYP1A1 3801T>C Polymorphism with Pathological Response to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer: A Cross-Sectional Study Using Residual Cancer Burden

Esther Sitorus
esther.reny@usu.ac.id
Universitas Sumatera Utara
Kamal Siregar
Universitas Sumatera Utara
Dedi Ardinata
Universitas Sumatera Utara
Lidya Laksmi
Universitas Sumatera Utara
Dedy Syahrizal
Universitas Syiah Kuala
Lili Rohmawati
Universitas Sumatera Utara
Mustafa Amin
Universitas Sumatera Utara
Muhammad Rusda
Universitas Sumatera Utara

Research Article

Keywords: Triple-negative breast cancer, CYP1A1, Polymorphism, Residual cancer burden, Neoadjuvant chemotherapy, Pharmacogenetics, Rs4646903

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11449430/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Association of CYP1A1 3801T>C Polymorphism with Pathological Response to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer: A Cross-Sectional Study Using Residual Cancer Burden

Abstract

Background Triple-negative breast cancer (TNBC) still depends on cytotoxic neoadjuvant chemotherapy, yet pathological response varies widely between patients. Cytochrome P450 1A1 is expressed in breast tissue and metabolises xenobiotics locally, and its 3801T>C (M1, rs4646903) variant is more frequent in Asian populations. Its relationship with neoadjuvant chemotherapy response in TNBC has not been reported.

Methods This analytic cross-sectional study aimed to test whether CYP1A1 3801T>C genotype is associated with pathological response, and included 98 women with TNBC treated with six cycles of neoadjuvant taxane–anthracycline chemotherapy followed by mastectomy at two tertiary referral hospitals in Indonesia between January 2022 and July 2025. CYP1A1 3801T>C was genotyped from formalin-fixed paraffin-embedded tumour DNA by polymerase chain reaction–restriction fragment length polymorphism using MspI. Pathological response was graded as Residual Cancer Burden (RCB) class 0 to III. Associations were tested with Fisher–Freeman–Halton exact tests and unadjusted proportional-odds ordinal logistic regression; clinicopathological predictors were assessed in a multivariable ordinal model.

Results Genotype frequencies were TT 32.7%, TC 49.0% and CC 18.4% (C allele 42.9%, 95% confidence interval 36.1–49.9), consistent with Hardy–Weinberg equilibrium ($p = 1.000$). RCB class was 0 in 3.1%, I in 8.2%, II in 58.2% and III in 30.6%. Genotype was not associated with RCB class ($p = 0.574$): the odds ratio was 0.57 (95% confidence interval 0.25–1.32) for TC+CC

versus TT, 0.60 (0.21–1.71) for CC versus TT+TC and 0.66 (0.37–1.17) per C allele. All three complete responders carried the C allele. Tumour size of 5 cm or more (odds ratio 4.29, 1.70–11.53) and lymphovascular invasion (odds ratio 4.68, 1.76–13.27) independently predicted a higher RCB class.

Conclusions CYP1A1 3801T>C was not significantly associated with pathological response in TNBC, although imprecise estimates leave a smaller effect possible. Markers of tumour burden dominated response in this cohort, and larger prospective studies with germline genotyping are needed.

Keywords Triple-negative breast cancer; CYP1A1; Polymorphism; Residual cancer burden; Neoadjuvant chemotherapy; Pharmacogenetics; Rs4646903

Background

Breast cancer is the most commonly diagnosed malignancy and the leading cause of cancer death among women worldwide [1]. GLOBOCAN 2020 estimated about 2.3 million new cases, or 11.7% of all cancers, which placed breast cancer ahead of lung cancer in incidence [1]. In Indonesia it is also the most frequent cancer in women, accounting for 16.6% of all cancers [2]. The incidence of breast cancer in younger women is also rising across Asia [27].

Breast cancer is heterogeneous in clinical presentation, tumour biology and response to chemotherapy [3, 5]. Triple-negative breast cancer (TNBC) is a distinct subtype defined by the absence of oestrogen receptor (ER) and progesterone receptor (PR) expression and of human epidermal growth factor receptor 2 (HER2) expression or amplification [3, 5]. It accounts for roughly 15–20% of breast cancers [3] and is characterised by high histological grade, early

recurrence and a poor prognosis [4, 5]. Because TNBC lacks the targets exploited by endocrine and anti-HER2 therapy, systemic treatment still relies largely on cytotoxic neoadjuvant chemotherapy followed by surgical resection and local therapy [6, 8].

The main difficulty is that response to neoadjuvant chemotherapy varies widely between patients. Some achieve a pathological complete response, whereas others have substantial residual disease and a worse outcome [9, 10]. There is also no universally accepted consensus on how pathological response should be assessed or predicted. Several scoring systems exist, including NSABP B-18, Miller–Payne and Residual Cancer Burden (RCB) [11, 12], and practice differs between countries and depends on local pathology consensus [11]. RCB combines residual tumour size, cellularity and nodal involvement into a continuous index and four classes. It has been validated as a prognostic measure and does not require pre-treatment tumour profiles [11, 12].

One candidate explanation for this variability is inter-individual difference in xenobiotic metabolism. Cytochrome P450 1A1 (CYP1A1) is a phase I enzyme that, unlike most cytochrome P450 enzymes, is expressed mainly outside the liver, including in breast tissue [14, 15]. It catalyses the oxidative phase I metabolism of xenobiotics [14, 16], and breast tumours can also convert steroid precursors locally [16]. Together, these observations raise the possibility that CYP1A1 metabolises xenobiotics, potentially including chemotherapeutic agents, in situ within the tumour and independently of hepatic metabolism. CYP1A1 genotype has been proposed as a predictor of chemotherapy effect in metastatic breast cancer treated with docetaxel-based regimens [21].

The CYP1A1 3801T>C (M1) polymorphism lies in the 3′-flanking region and creates an MspI restriction site. It has been linked to increased carcinogen activation [16, 18]. The variant is

considerably more frequent in Asian populations; in a pooled analysis the homozygous CC genotype was found in about 13% of Asians versus 1% of Whites and 6% of Blacks [14, 18]. Studies of its association with breast cancer risk have reported increased, decreased and no association [19, 20], so its functional consequences are not yet completely defined.

To our knowledge, no study has evaluated the relationship between CYP1A1 3801T>C and the pathological response of TNBC to neoadjuvant chemotherapy. Accordingly, this study aimed to analyse the association between CYP1A1 3801T>C genotype and pathological response to neoadjuvant chemotherapy, assessed by RCB class, in patients with TNBC. We hypothesised that genotype is associated with RCB class. Secondary objectives were to describe genotype and allele distributions and their conformity to Hardy–Weinberg equilibrium, and to examine the associations of genotype, clinicopathological features and p53 and Ki-67 expression with RCB.

Methods

Study design and setting

This was an analytic cross-sectional study using archived formalin-fixed paraffin-embedded (FFPE) tumour blocks and medical records of patients with TNBC treated at two tertiary referral hospitals in Indonesia between January 2022 and July 2025. Histopathological review was performed at the departments of anatomical pathology of both hospitals and at the histology laboratory of the authors' institution. Molecular analysis was performed at the integrated laboratory of the same institution. The study is reported according to the STROBE statement for cross-sectional studies.

Participants

Patients were enrolled by consecutive sampling. Eligible patients had histologically confirmed invasive carcinoma of no special type (IC-NST) or another histological type with a recorded

histological grade, a triple-negative immunophenotype (ER−, PR−, HER2−) with p53 and Ki-67 results available, and complete neoadjuvant chemotherapy with a taxane and an anthracycline (six cycles) followed by mastectomy. Pre-treatment (diagnostic biopsy) and post-treatment (mastectomy) FFPE blocks were collected for each patient. Patients were excluded if either block was unavailable, if a block had been stored for more than 5 years, or if a block was unsuitable after validation (inadequate haematoxylin–eosin staining, immunohistochemistry or DNA quality). Patients whose DNA could not be genotyped were also excluded. Of 100 patients enrolled, two were excluded because genotyping failed, leaving 98 for analysis.

Data sources and measurement

Histopathology. Tissue was fixed in 10% neutral-buffered formalin, embedded in paraffin and stained with haematoxylin–eosin. Diagnosis and grading were made by the investigator and confirmed by two pathologists.

Immunohistochemistry. Four-micrometre FFPE sections were stained on a Leica BOND automated immunostainer using the following antibodies (Leica Biosystems): ER (clone 6F11, REF PA0151), PR (clone 16, REF PA0312), HER2 (Novocastra NCL-L-CB11), p53 (clone DO-7, REF PA0057) and Ki-67 (clone MM1, REF PA0118). Pre-treatment biopsy sections were used for ER, PR, HER2, p53 and Ki-67. ER and PR were considered negative when nuclear staining was present in 0% of tumour cells, and HER2 was considered negative at score 0 or 1+, or 2+ with negative in situ hybridisation. p53 expression was scored as 0, 1+, 2+ or 3+, and Ki-67 was expressed as the percentage of positively stained tumour nuclei.

Genotyping. DNA was extracted from FFPE blocks using the Wizard Genomic DNA Purification Kit (Promega). The CYP1A1 3801T>C region was amplified using the GoTaq Core System I (Promega) on an Applied Biosystems Veriti 96-well thermal cycler with the primers 5′-

CAGTGAAGAGGTGTAGCCGCT-3′ (forward) and 5′-TAGGAGTCTTGTCTCATGCCT-3′ (reverse). The cycling conditions were 94 °C for 5 min, then 30 cycles of 94 °C for 1 min, 60 °C for 60 s and 72 °C for 60 s, followed by a final extension at 72 °C for 10 min. Products were digested with MspI, which cuts at the site created by the C allele. Fragments were separated on a 4% agarose gel stained with [stain] and visualised under ultraviolet transillumination (Uvitec Cambridge) alongside a 100 bp DNA ladder. Genotypes were assigned as TT (340 bp), TC (340, 200 and 140 bp) or CC (200 and 140 bp). Complete digestion was confirmed [please state how: by a known-genotype control sample and/or by repeat digestion].

Pathological response. RCB was calculated from the post-mastectomy specimen using the MD Anderson RCB calculator [11, 13]. Inputs were the two largest dimensions of the residual tumour bed, overall cancer cellularity, the percentage of in situ carcinoma, the number of positive lymph nodes and the diameter of the largest nodal metastasis [11, 12]. Cases were assigned to RCB-0, I, II or III.

Bias

Patients were enrolled consecutively and the same criteria were applied at both hospitals to limit selection bias. Only patients who completed six cycles of neoadjuvant chemotherapy, underwent mastectomy and had suitable archived blocks were analysed, so patients who stopped treatment early or whose blocks were unusable are not represented. Blocks stored for more than 5 years were excluded to reduce DNA degradation, and block suitability was checked before analysis. Diagnosis was confirmed by pathologists and RCB was calculated with the standard calculator to reduce measurement variability, but inter-observer variability in p53, Ki-67, tumour-infiltrating lymphocyte (TIL) and lymphovascular invasion (LVI) scoring cannot be excluded. Genotyping and RCB assessment were blinded. Genotype is determined by germline sequence and does not

change with treatment. All patients received a taxane–anthracycline regimen for six cycles, which limits treatment variation; the association between genotype and RCB is reported unadjusted, and residual confounding by unmeasured factors such as clinical stage, specific agents, doses and dose intensity is possible. Ethnicity was examined descriptively against genotype.

Study size

The minimum sample size was calculated with a two-proportion formula ($\alpha = 0.05$, power 80%), yielding 45 participants (based on the proportion of grade 3 tumours, 73.9% in Ibrahiem et al. [22], versus 33.3% in tumours with a low TIL level). Because reliable estimation of genotype frequencies and testing of Hardy–Weinberg equilibrium require a larger sample than this minimum, the target was set at 100 patients, recruited by consecutive sampling. Two patients were excluded after validation because of failed electrophoresis, leaving 98 patients for analysis.

Statistical analysis

Categorical variables are presented as frequencies and percentages. Associations between genotype and patient characteristics were tested with Fisher's exact test (Fisher–Freeman–Halton exact test for tables larger than 2 × 2) under the genotypic, dominant and recessive models. Hardy–Weinberg equilibrium was assessed with the chi-square test, and allele frequencies are given with Wilson 95% confidence intervals (CIs). For analysis, Ki-67 was dichotomised as low (<30%) or high (≥30%), p53 was treated as an ordinal score (0 to 3+), and TILs were categorised as low (<10%), moderate (10–40%) or high (>40%).

The association between genotype and RCB class was tested with the Fisher–Freeman–Halton exact test and estimated with unadjusted proportional-odds ordinal logistic regression (additive, dominant, recessive and genotypic models), reporting odds ratios (ORs) with 95% CIs. RCB-III

versus RCB-0 to II was compared with Fisher's exact test, with Woolf confidence intervals for the odds ratio. Factors associated with RCB class (age, histological type, tumour size, LVI, TILs, histological grade, p53 expression and Ki-67) were examined in a multivariable ordinal logistic regression, reporting odds ratios with 95% CIs, likelihood-ratio tests and McFadden R^2. Analyses were performed in jamovi 2.6. Two-sided $p < 0.05$ was considered significant; as the analyses were exploratory, no correction for multiple testing was applied.

Ethical approval

The study was approved by the health research ethics committee of the authors' institution [approval number withheld for anonymised review]. Permission from both participating hospitals was obtained before committee review, so the clearance covers both sites. The study used archived tissue blocks and anonymised medical records.

Results

Participants and patient characteristics

One hundred patients were enrolled. Two were excluded because genotyping failed, leaving 98 patients for analysis. Most patients were aged 50 years or younger (54/98, 55.1%), and the largest ethnic groups were Batak (47, 48.0%) and Javanese (33, 33.7%); 65 patients (66.3%) were housewives. Nearly all tumours were IC-NST (91, 92.9%) and histological grade 3 (93, 94.9%). Tumour category was T1 in 20 (20.4%), T2 in 52 (53.1%) and T3 in 26 patients (26.5%). LVI was present in 30 patients (30.6%), TILs were low in 62 (63.3%), and Ki-67 was high (≥30%) in 94 (95.9%). The distribution of p53 expression scores was 0 in 29 (29.6%), 1+ in 20 (20.4%), 2+ in 16 (16.3%) and 3+ in 33 patients (33.7%). Characteristics by genotype are shown in Table 1.

CYP1A1 3801T>C genotype distribution

The genotype frequencies were TT 32 (32.7%), TC 48 (49.0%) and CC 18 (18.4%), corresponding to a T allele frequency of 57.1% (112/196) and a C allele frequency of 42.9% (84/196) (Table 2, Fig. 1). The genotype distribution was in Hardy–Weinberg equilibrium (χ^2 = 0.000, df = 1, p = 1.000). A representative PCR–RFLP gel is shown in Fig. 2.

Genotype and other patient characteristics

Genotype was not associated with age, ethnicity, histological type, grade, T category, tumour size, LVI, TILs, p53 expression or Ki-67 under the three-genotype comparison (all p > 0.05; Table 1). The dominant (TC+CC versus TT) and recessive (CC versus TT+TC) models gave the same result (Supplementary Table S1). All 48 patients with the TC genotype had a high Ki-67 index, compared with 29 of 32 TT and 17 of 18 CC patients (p = 0.075;).

Genotype and pathological response

The RCB distribution in the whole cohort was RCB-0 (pathological complete response) in 3 patients (3.1%; 95% CI 1.0–8.6), RCB-I in 8 (8.2%), RCB-II in 57 (58.2%) and RCB-III in 30 patients (30.6%; 95% CI 22.4–40.3). All three patients with RCB-0 carried the C allele (two TC, one CC); none had the TT genotype.

RCB class was not significantly associated with genotype overall (p = 0.574), or under the dominant (p = 0.625) or recessive (p = 0.300) model (Table 3). RCB-III occurred in 37.5% of TT, 27.1% of TC and 27.8% of CC patients.

In unadjusted ordinal logistic regression, the odds of a higher (worse) RCB class were not significantly different by genotype: OR 0.57 (95% CI 0.25–1.32) for TC+CC versus TT, 0.60 (0.21–1.71) for CC versus TT+TC and 0.66 (0.37–1.17) per C allele (Table 4). Results were similar when RCB was dichotomised as RCB-III versus RCB-0 to II (Table 4).

Clinicopathological factors and RCB class

In the multivariable ordinal logistic regression of clinicopathological factors and biomarkers (n = 98; model $\chi^2(9) = 20.2$, $p = 0.017$; McFadden $R^2 = 0.104$), a tumour size of 5 cm or more (OR 4.29, 95% CI 1.70–11.53; p = 0.003) and the presence of LVI (OR 4.68, 95% CI 1.76–13.27; p = 0.003) were associated with a higher RCB class. Age, histological type, grade, TILs, p53 expression and Ki-67 were not significantly associated with RCB class (Table 5). Estimates for histological type, grade and Ki-67 were imprecise because of the small number of patients in the reference or comparison categories.

Discussion

In this cross-sectional study of 98 patients with TNBC treated with neoadjuvant taxane–anthracycline chemotherapy at two hospitals, CYP1A1 3801T>C genotype was not significantly associated with RCB class (exact p = 0.574; dominant model OR 0.57, 95% CI 0.25–1.32; recessive model OR 0.60, 95% CI 0.21–1.71). Only three patients (3.1%) achieved a pathological complete response, and all three carried the C allele. Tumour size of 5 cm or more and LVI were associated with a worse RCB class. To our knowledge, this is the first study to examine this polymorphism in relation to neoadjuvant chemotherapy response in TNBC.

The C allele frequency was 42.9% (95% CI 36.1–49.9), in line with the higher frequencies reported in Asian than in European populations [19]. The genotype distribution was in Hardy–Weinberg equilibrium, which supports the reliability of the genotyping. Because the analysis was restricted to patients with cancer, however, equilibrium does not exclude selection or population stratification, and we do not interpret it further.

The point estimates for the association between genotype and RCB favoured C allele carriers, with lower odds of a higher RCB class, but the confidence intervals were wide and included both

a moderate benefit and a modest harm. With 98 patients and only three in RCB-0, we estimate that the study could detect an odds ratio of roughly 0.3 or less, or 3.3 or more, with 80% power. The findings therefore neither confirm nor exclude a smaller effect, and the descriptive observation that all complete responders were carriers should be treated as hypothesis-generating. The closest prior evidence comes from a different setting: Zhou et al. [21] proposed CYP1A1 genotype as a predictor of chemotherapy effect in metastatic breast cancer treated with docetaxel-based regimens, rather than in TNBC or in the neoadjuvant setting.

Several factors may explain the absence of an association. First, the principal known pathways of the drugs used here do not involve CYP1A1: taxanes are metabolised mainly by CYP2C8 and CYP3A4/5 [23], and doxorubicin is inactivated mainly by carbonyl reduction through CBR1 and AKR1C3 [24, 25]. A contribution of CYP1A1 would therefore have to come from minor or tissue-specific pathways. Second, 3801T>C lies in the 3′-flanking region and its functional effect is not completely defined, while CYP1A1 expression is inducible by environmental ligands, such as polycyclic aromatic hydrocarbons, through the aryl hydrocarbon receptor [16, 17]; genotype may therefore not reflect enzyme activity within the tumour. Third, DNA was extracted from tumour blocks, which may differ from germline DNA. Finally, response in this cohort appears to be driven largely by tumour burden, which may mask subtle pharmacogenetic effects.

The pathological response in this cohort was poor: 88.8% of patients had RCB-II or III, and the pathological complete response rate of 3.1% is well below the roughly 30% reported in pooled analyses of anthracycline–taxane regimens [9] and below the 15.2% reported in an Asian series [26]. Possible contributors are the high tumour burden at presentation (41.8% of tumours ≥5 cm, 26.5% T3), LVI in 30.6%, low TIL levels in 63.3%, and treatment limited to a taxane and an anthracycline without carboplatin or immunotherapy. RCB is a validated prognostic measure

across breast cancer phenotypes [10], so this distribution suggests a substantial residual disease burden in this population.

The clinicopathological profile was typical of TNBC: more than half of the patients were aged 50 years or younger [27], and almost all tumours were IC-NST and grade 3 [3]. Genotype was not associated with age, ethnicity, histological type, grade, tumour size, LVI, TILs or p53 expression. Ki-67 was high in 95.9% of patients, and all 48 patients with the TC genotype had a high index ($p = 0.075$). Because only four patients had a low Ki-67, this reflects sparse data rather than a biological signal.

In the multivariable analysis, tumour size of 5 cm or more (OR 4.29, 95% CI 1.70–11.53) and LVI (OR 4.68, 95% CI 1.76–13.27) were associated with a worse RCB class. This agrees with reports that larger tumours are less likely to reach a complete response [12] and that LVI marks an adverse prognosis [28]. We did not find an association of TILs, p53 expression or Ki-67 with RCB class, unlike studies in which higher TIL levels predicted pathological complete response. Differences in TIL cut-offs (<10%, 10–40% and >40% here, against the 30% or 50% thresholds recommended by the International TILs Working Group [29]), the limited sample and the sparse categories may explain this. These estimates come from an exploratory model with nine parameters in 98 patients and should be interpreted cautiously.

These findings have modest clinical implications. Tumour size and LVI could help identify patients unlikely to respond to standard chemotherapy, and the low pathological complete response rate supports evaluating treatment intensification, such as the addition of pembrolizumab, which increased the pathological complete response rate from 51.2% to 64.8% in KEYNOTE-522 [30], recognising that access to these options is limited in Indonesia. Reporting RCB routinely would support such decisions. Future work should include larger

prospective multicentre cohorts, germline DNA, detailed records of agents and doses, CYP1A1 expression or activity, additional pharmacogenetic variants and survival outcomes.

Strengths of this study include recruitment from two hospitals, standardised RCB assessment, a genotyping failure rate of only 2% and genotype frequencies in Hardy–Weinberg equilibrium.

The study also has several limitations. The cross-sectional design precludes conclusions about survival or causality. The sample was small, with wide confidence intervals and only three RCB-0 cases, and the analyses were exploratory with no correction for multiple testing. The association between genotype and RCB was not adjusted for other factors. Only one polymorphism was studied, and DNA came from tumour rather than germline tissue. Clinical stage, specific agents and doses were not available. TILs, p53 and Ki-67 were scored manually, so inter-observer variability is possible, and some categories were sparse (four patients with a low Ki-67, five with grade 2 tumours). Finally, all patients came from one region of Indonesia, which may limit generalisability.

Conclusions

In this cohort of patients with TNBC, CYP1A1 3801T>C was not significantly associated with pathological response to neoadjuvant taxane–anthracycline chemotherapy, although the estimates were imprecise and a smaller effect cannot be excluded. Larger tumour size and LVI were associated with a worse RCB class, which supports their use in identifying patients unlikely to respond to standard chemotherapy. Larger prospective studies with germline genotyping and detailed treatment data are needed.

List of abbreviations

AhR: Aryl hydrocarbon receptor

AKR1C3: Aldo-keto reductase family 1 member C3

CBR1: Carbonyl reductase 1

CI: Confidence interval

CYP1A1: Cytochrome P450 family 1 subfamily A member 1

DNA: Deoxyribonucleic acid

ER: Oestrogen receptor

FFPE: Formalin-fixed paraffin-embedded

HER2: Human epidermal growth factor receptor 2

IC-NST: Invasive carcinoma of no special type

LVI: Lymphovascular invasion

OR: Odds ratio

pCR: Pathological complete response

PCR–RFLP: Polymerase chain reaction–restriction fragment length polymorphism

PR: Progesterone receptor

RCB: Residual cancer burden

SNP: Single nucleotide polymorphism

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

TILs: Tumour-infiltrating lymphocytes

TNBC: Triple-negative breast cancer

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209-49.

2. International Agency for Research on Cancer. Global Cancer Observatory: Indonesia fact sheet, GLOBOCAN 2020. https://gco.iarc.fr/today. Accessed 20 Oct 2026.

3. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. N Engl J Med. 2010;363:1938-48.

4. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res. 2007;13:4429-34.

5. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol. 2016;13:674-90.

6. Won KA, Spruck C. Triple-negative breast cancer therapy: current and future perspectives. Int J Oncol. 2020;57:1245-61.

7. van den Ende NS, Nguyen AH, Jager A, Kok M, Debets R, van Deurzen CHM. Triple-negative breast cancer and predictive markers of response to neoadjuvant chemotherapy: a systematic review. Int J Mol Sci. 2023;24:2969.

8. Gradishar WJ, Moran MS, Abraham J, Aft R, Agnese D, Allison KH, et al. Breast cancer, version 4.2023: featured updates to the NCCN guidelines. J Natl Compr Canc Netw. 2023;21:594-608.

9. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet. 2014;384:164-72.

10. Symmans WF, Wei C, Gould R, Yu X, Zhang Y, Liu M, et al. Long-term prognostic risk after neoadjuvant chemotherapy associated with residual cancer burden and breast cancer subtype. J Clin Oncol. 2017;35:1049-60.

11. Symmans WF, Peintinger F, Hatzis C, Rajan R, Kuerer H, Valero V, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. J Clin Oncol. 2007;25:4414-22.

12. Hamy AS, Darrigues L, Laas E, De Croze D, Topciu L, Lam GT, et al. Prognostic value of the residual cancer burden index according to breast cancer subtype: validation on a cohort of BC patients treated by neoadjuvant chemotherapy. PLoS One. 2020;15:e0234191.

13. MD Anderson Cancer Center. Residual cancer burden calculator: detailed pathology methods for using residual cancer burden. https://www3.mdanderson.org/app/medcalc/index.cfm?pagename=jsconvert3. Accessed 20 Oct 2026.

14. Androutsopoulos VP, Tsatsakis AM, Spandidos DA. Cytochrome P450 CYP1A1: wider roles in cancer progression and prevention. BMC Cancer. 2009;9:187.

15. Rodriguez M, Potter DA. CYP1A1 regulates breast cancer proliferation and survival. Mol Cancer Res. 2013;11:780-92.

16. Liu D, Wu D, Zhao L, Yu A, Zhang Y, Wang Y, et al. Single nucleotide polymorphisms distant from xenobiotic response elements can modulate aryl hydrocarbon receptor function: SNP-dependent CYP1A1 induction. Drug Metab Dispos. 2018;46:1372-81.

17. Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. J Biol Chem. 2004;279:23847-50.

18. Masson LF, Sharp L, Cotton SC, Little J. Cytochrome P-450 1A1 gene polymorphisms and risk of breast cancer: a HuGE review. Am J Epidemiol. 2005;161:901-15.

19. Qin Q, Tan QW, Zhang R, Mo QG, Lu J, Liu DF, et al. Association between CYP1A1 polymorphisms and cancer susceptibility: a meta-analysis based on 268 case-control studies. J Cancer. 2014;5:703-18.

20. Yang C, He XF. Evaluation of association studies and a systematic review and meta-analysis of CYP1A1 T3801C and A2455G polymorphisms in breast cancer risk. PLoS One. 2021;16:e0249632.

21. Zhou X, Qiao G, Wang X, Song Q, Morse MA, Hobeika A, et al. CYP1A1 genetic polymorphism is a promising predictor to improve chemotherapy effects in patients with metastatic breast cancer treated with docetaxel plus thiotepa vs. docetaxel plus capecitabine. Cancer Chemother Pharmacol. 2018;81:365-72.

22. Ibrahiem AT, Mouhamed HA, Anter AH. Predictive significance of stromal-associated lymphocytes in response to neoadjuvant chemotherapy in breast cancer. Menoufia Med J. 2017;30:1168-77.

23. de Weger VA, Beijnen JH, Schellens JH. Cellular and clinical pharmacology of the taxanes docetaxel and paclitaxel: a review. Anticancer Drugs. 2014;25:488-94.

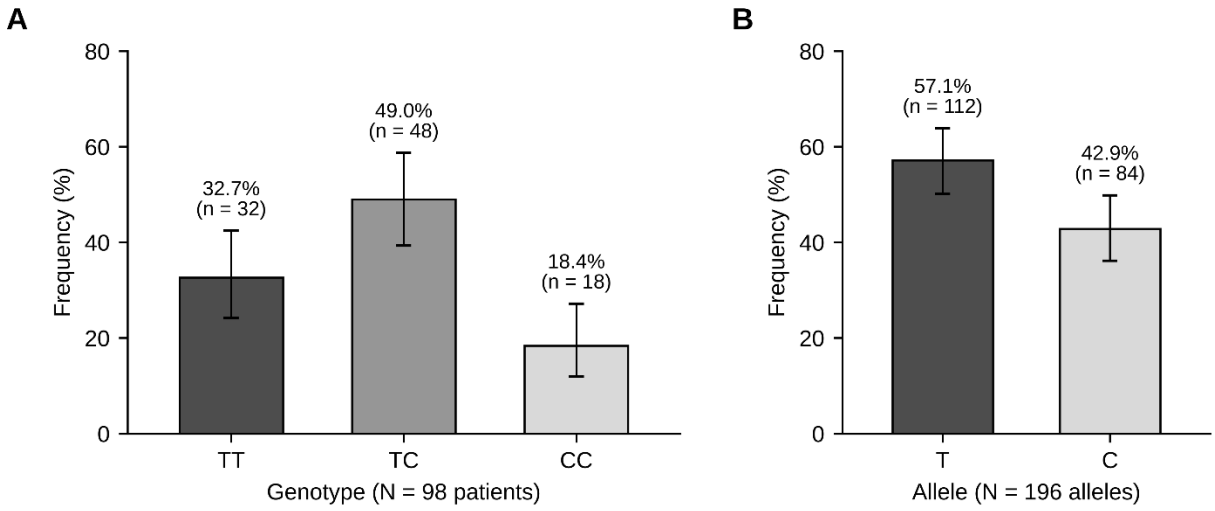
24. Kassner N, Huse K, Martin HJ, Godtel-Armbrust U, Metzger A, Meineke I, et al. Carbonyl reductase 1 is a predominant doxorubicin reductase in the human liver. Drug Metab Dispos. 2008;36:2113-20.

25. Novotna R, Wsol V, Xiong G, Maser E. Inactivation of the anticancer drugs doxorubicin and oracin by aldo-keto reductase (AKR) 1C3. Toxicol Lett. 2008;181:1-6.

26. Xu X, Zhao W, Liu C, Gao Y, Chen D, Wu M, et al. The residual cancer burden index as a valid prognostic indicator in breast cancer after neoadjuvant chemotherapy. BMC Cancer. 2024;24:13.

389 27. Wang C, Kar S, Lai X, Cai W, Arfuso F, Sethi G, et al. Triple negative breast cancer in Asia:
390 an insider's view. Cancer Treat Rev. 2018;62:29-38.
391 28. Hamy AS, Lam GT, Laas E, Darrigues L, Balezeau T, Guerin J, et al. Lymphovascular
392 invasion after neoadjuvant chemotherapy is strongly associated with poor prognosis in breast
393 carcinoma. Breast Cancer Res Treat. 2018;169:295-304.
394 29. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation
395 of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International
396 TILs Working Group 2014. Ann Oncol. 2015;26:259-71.
397 30. Schmid P, Cortes J, Pusztai L, McArthur H, Kummel S, Bergh J, et al. Pembrolizumab for
398 early triple-negative breast cancer. N Engl J Med. 2020;382:810-21.

399 **Figure legends**



400

401 **Fig. 1** Genotype and allele frequencies of the CYP1A1 3801T>C (rs4646903) polymorphism in
402 patients with triple-negative breast cancer (N = 98). **a** Genotype distribution: TT (wild type)
403 32.7% (n = 32); TC (heterozygous) 49.0% (n = 48); CC (homozygous variant) 18.4% (n = 18). **b**
404 Allele distribution (2N = 196 alleles): T 57.1% (n = 112); C 42.9% (n = 84). Bars show

frequencies (%) and error bars show Wilson 95% confidence intervals. Genotype frequencies were in Hardy–Weinberg equilibrium ($\chi^2 = 0.000$, df = 1, p = 1.000). Genotyping was performed by PCR–RFLP (MspI) on DNA from formalin-fixed paraffin-embedded tumour tissue.

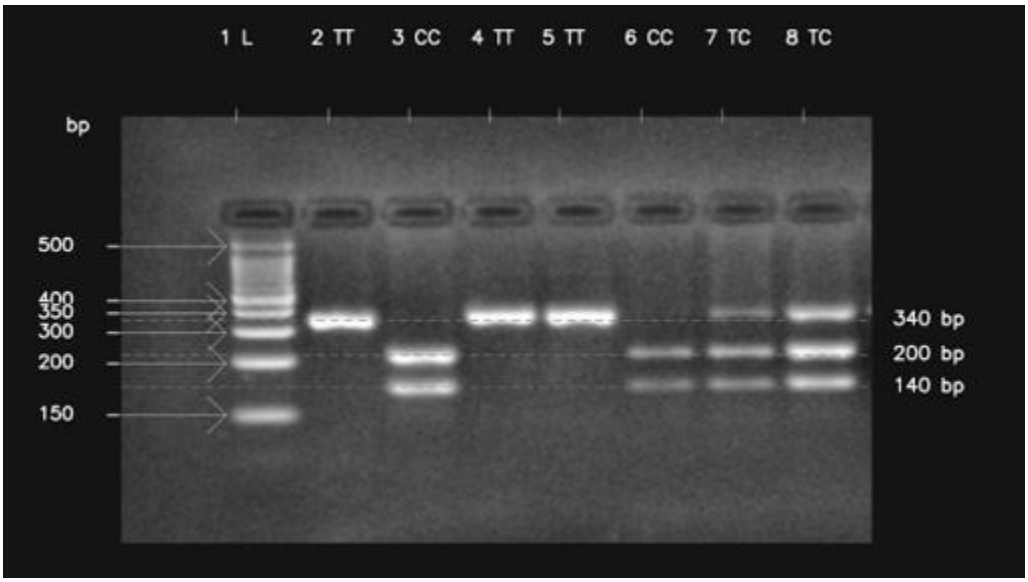


Fig. 2 Representative PCR–RFLP analysis of the CYP1A1 3801T>C (rs4646903) polymorphism. The 340 bp PCR product was digested with MspI, which cuts only the C allele, and the fragments were separated on a 4% agarose gel stained with [stain] and visualised under ultraviolet transillumination (Uvitec Cambridge). Lane 1 (L), 100 bp DNA ladder; lanes 2, 4 and 5, TT homozygous wild type (undigested 340 bp); lanes 3 and 6, CC homozygous variant (200 and 140 bp); lanes 7 and 8, TC heterozygous (340, 200 and 140 bp). bp, base pairs.

Tables

Table 1 Characteristics of the patients, overall and by CYP1A1 3801T>C genotype

Characteristic	Total (n = 98)	TT (n = 32)	TC (n = 48)	CC (n = 18)	p*
Age, years					
≤50	54 (55.1)	17 (53.1)	27 (56.2)	10 (55.6)	0.963
>50	44 (44.9)	15 (46.9)	21 (43.8)	8 (44.4)	
Ethnicity					
Batak	47 (48.0)	16 (50.0)	27 (56.2)	4 (22.2)	0.166
Javanese	33 (33.7)	8 (25.0)	14 (29.2)	11 (61.1)	

Characteristic	Total (n = 98)	TT (n = 32)	TC (n = 48)	CC (n = 18)	p*
Others†	18 (18.4)	8 (25.0)	7 (14.6)	3 (16.7)	
Histological type					
IC-NST	91 (92.9)	30 (93.8)	45 (93.8)	16 (88.9)	0.763
Non-NST	7 (7.1)	2 (6.2)	3 (6.2)	2 (11.1)	
Histological grade					
2	5 (5.1)	0 (0.0)	3 (6.2)	2 (11.1)	0.155
3	93 (94.9)	32 (100.0)	45 (93.8)	16 (88.9)	
T category					
T1 (≤2 cm)	20 (20.4)	4 (12.5)	10 (20.8)	6 (33.3)	0.440
T2 (>2–5 cm)	52 (53.1)	17 (53.1)	26 (54.2)	9 (50.0)	
T3 (>5 cm)	26 (26.5)	11 (34.4)	12 (25.0)	3 (16.7)	
Tumour size					
<5 cm	57 (58.2)	16 (50.0)	30 (62.5)	11 (61.1)	0.542
≥5 cm‡	41 (41.8)	16 (50.0)	18 (37.5)	7 (38.9)	
Lymphovascular invasion					
Present	30 (30.6)	12 (37.5)	12 (25.0)	6 (33.3)	0.494
Absent	68 (69.4)	20 (62.5)	36 (75.0)	12 (66.7)	
TILs					
Low	62 (63.3)	18 (56.2)	34 (70.8)	10 (55.6)	0.521
Moderate	10 (10.2)	3 (9.4)	4 (8.3)	3 (16.7)	
High	26 (26.5)	11 (34.4)	10 (20.8)	5 (27.8)	
p53 expression score					
0	29 (29.6)	7 (21.9)	16 (33.3)	6 (33.3)	0.827
1+	20 (20.4)	8 (25.0)	8 (16.7)	4 (22.2)	
2+	16 (16.3)	4 (12.5)	9 (18.8)	3 (16.7)	
3+	33 (33.7)	13 (40.6)	15 (31.2)	5 (27.8)	
Ki-67 index					
Low (<30%)	4 (4.1)	3 (9.4)	0 (0.0)	1 (5.6)	0.075
High (≥30%)	94 (95.9)	29 (90.6)	48 (100.0)	17 (94.4)	

*Values are n (%); percentages are column percentages. *Fisher's exact test (Fisher–Freeman–Halton exact test for tables larger than 2 × 2), three-genotype comparison. †Melayu (5), Minang (4), Chinese (3), Acehnese (2), Gayo (2), Nias (2); the p value for ethnicity was calculated using the eight original categories. ‡Includes tumours measuring exactly 5 cm. IC-NST invasive carcinoma of no special type, TILs tumour-infiltrating lymphocytes*

Table 2 Genotype and allele frequencies of CYP1A1 3801T>C and Hardy–Weinberg equilibrium (n = 98)

	n	% (95% CI)
Genotype		
TT (wild type)	32	32.7 (24.2–42.4)
TC (heterozygous)	48	49.0 (39.3–58.7)
CC (homozygous variant)	18	18.4 (11.9–27.2)
Allele (2n = 196)		
T	112	57.1 (50.1–63.9)
C	84	42.9 (36.1–49.9)
Hardy–Weinberg equilibrium	$\chi^2 = 0.000$	df = 1, p = 1.000

Confidence intervals are Wilson 95% intervals. CI confidence interval

Table 3 CYP1A1 3801T>C genotype and RCB class

Model / genotype	RCB-0	RCB-I	RCB-II	RCB-III	Total	p†
Genotypic						0.574
TT	0 (0.0)	2 (6.2)	18 (56.2)	12 (37.5)	32	
TC	2 (4.2)	3 (6.2)	30 (62.5)	13 (27.1)	48	
CC	1 (5.6)	3 (16.7)	9 (50.0)	5 (27.8)	18	
Dominant						0.625
TT	0 (0.0)	2 (6.2)	18 (56.2)	12 (37.5)	32	
TC+CC	3 (4.5)	6 (9.1)	39 (59.1)	18 (27.3)	66	
Recessive						0.300
TT+TC	2 (2.5)	5 (6.2)	48 (60.0)	25 (31.2)	80	
CC	1 (5.6)	3 (16.7)	9 (50.0)	5 (27.8)	18	
All patients	3 (3.1)	8 (8.2)	57 (58.2)	30 (30.6)	98	

Values are n (%); percentages are row percentages. †Fisher–Freeman–Halton exact test. RCB residual cancer burden; RCB-0 pathological complete response, RCB-I minimal residual burden, RCB-II moderate residual burden, RCB-III extensive residual burden

Table 4 Unadjusted association between CYP1A1 3801T>C genotype and RCB class

Genetic model	Comparison	RCB class (ordinal), OR (95% CI)	p	RCB-III vs RCB-0 to II, OR (95% CI)	p
Additive	Per C allele	0.66 (0.37–1.17)	0.156	–	–
Dominant	TC+CC vs TT	0.57 (0.25–1.32)	0.189	0.62 (0.25–1.53)	0.353
Recessive	CC vs TT+TC	0.60 (0.21–1.71)	0.337	0.85 (0.27–2.63)	1.000
Genotypic	TC vs TT	0.62 (0.26–1.49)	0.287	–	–
Genotypic	CC vs TT	0.45 (0.14–1.46)	0.183	–	–

An OR below 1 indicates lower odds of a higher RCB class (better response). Ordinal estimates are from proportional-odds logistic regression (Wald p); the dichotomised comparison uses

Fisher's exact test, with Woolf 95% confidence intervals for the odds ratio. No adjustment for other variables. CI confidence interval, OR odds ratio, RCB residual cancer burden

Table 5 Clinicopathological factors and biomarkers associated with RCB class (multivariable ordinal logistic regression, n = 98)

Variable	Comparison	OR	95% CI	p
Age	>50 vs ≤50 years	1.43	0.59–3.51	0.432
Histological type	Non-NST vs IC-NST	3.79	0.66–23.78	0.139
Tumour size	≥5 vs <5 cm	4.29	1.70–11.53	0.003
Lymphovascular invasion	Present vs absent	4.68	1.76–13.27	0.003
TILs	Moderate vs low	0.94	0.23–3.82	0.931
TILs	High vs low	1.19	0.45–3.21	0.725
Histological grade	3 vs 2	2.18	0.31–15.81	0.434
p53 expression	Per score increase	0.88	0.61–1.25	0.476
Ki-67	High vs low	0.38	0.02–4.18	0.460

An OR above 1 indicates higher odds of a higher (worse) RCB class. Model $\chi^2(9) = 20.2$, $p = 0.017$; McFadden $R^2 = 0.104$. Likelihood-ratio tests: tumour size $p = 0.002$; LVI $p = 0.002$. CI confidence interval, IC-NST invasive carcinoma of no special type, LVI lymphovascular invasion, OR odds ratio, RCB residual cancer burden, TILs tumour-infiltrating lymphocytes

Supplementary information

Supplementary Table S1 Association of CYP1A1 3801T>C genotype with patient characteristics under three genetic models (Fisher's exact test, p values)

Characteristic	Genotypic (TT / TC / CC)	Dominant (TC+CC vs TT)	Recessive (CC vs TT+TC)
Age (≤50 vs >50 years)	0.963	0.831	1.000
Ethnicity	0.166	0.293	0.087
Histological type	0.763	1.000	0.609
Histological grade	0.155	0.170	0.227
T category	0.440	0.286	0.302
Tumour size (<5 vs ≥5 cm)	0.542	0.281	1.000
Lymphovascular invasion	0.494	0.353	0.783
TILs	0.521	0.518	0.524
p53 expression score	0.827	0.473	0.951
Ki-67 (low vs high)	0.075	0.101	0.562

448 *TILs tumour-infiltrating lymphocytes*